arbitrary audio signal masking a complicated type of coding distortion, the JND description is by necessity approximate and empirical.

The result of Fig. 11(a) is one such empirical description, for the example of a trumpet signal with a bandwidth of 16 kHz. In this example, the staircase description of the JND profile is the result of interpolating between known results of noise-making-tone and tone-masking-noise [66]. The interpolation algorithm is based on a measure of how tone-like or noise-like the input signal is. The staircase treads in Fig. 11(a) correspond to the 25 *critical bands* in the human auditory system, narrow at low frequencies and wide at the higher frequencies. The phenomenon of masking is largely localized in each critical band, although there is some degree of interband masking (masking of distortion in a band by a signal in a different critical band). The interband masking effect falls at about 15 dB per critical band for the higher masked frequencies, and at about 25 dB per critical band for the lower masked frequencies. This behavior is related to the frequency shape of the cochlear filter.

The JND is a function of a (finely described) input spectrum and the ear model, rather than a simple transformation of the LPC spectrum of the input signal. While it is conceivable that a very high-order LPC analysis can describe the input spectrum finely enough to permit a useful JND model, current methodologies for transparent or near-transparent audio coding have depended on high-resolution frequency analysis using either a subband coding or a transform-coding framework.

Fig. 11(b) illustrates the typical characteristics of a 96-tap *quadrature-mirror filterbank* (QMF) used for subband analysis of audio, and the alternative framework of a 512-line *modified discrete cosine transform* (MDCT).

The QMF system is based on the division of a frequency band into contiguous but overlapping subbands. The partially-completed broken line shows the characteristic of a highpass filter that is the mirror image of the solid-line lowpass filter characteristic. The extent of overlap is a decreasing function of the number of filter taps. But the allowing of a nonzero overlap simplifies filter design. Frequency aliasing is caused in QMF analysis by sampling each of the two bands in the QMF split at twice the *nominal* bandwidth (rather than twice the actual, say 90 dB, bandwidth). However, with a special design of QMF filters, the process of QMF synthesis provides cancellation of this aliasing if the quantization noise inserted in the system is zero [16], [26], [62], [65], [140].

The modified DCT system [111] is a dual of the QMF approach in that it permits an overlap between successive transform blocks in the time-domain but decimates the resulting sequence to maintain the original sampling rate. In place of the frequency aliasing in the QMF system, the MDCT exhibits time-domain aliasing. But this is canceled by the inverse MDCT process in the receiver due to the design of the DCT basis vectors and the analysis window. The overlap in the MDCT is typically 50% and the decimation rate is 2:1 [97],[111]. While a 50% frequency-domain overlap is admissible in QMF design in principle, it is untypical.

Although QMF and MDCT systems are essentially dual, they exhibit different properties in the context of a specific overall coder. For example, operations such as signal anticipation and temporal bit allocation can be used to some extent to control quantizing distortion and the consequent phenomenon of uncanceled time-domain aliasing in the MDCT system. This, in turn, permits the use of a 50% time overlap which provides a smooth handling of the time process as well as a simple decimator design. In QMF design, on the other hand, one still prefers, in the current state of the art, to employ a small spectral overlap within the constraints on filter complexity. This, in turn, implies a relatively discontinuous handling of the frequency process. Finally, in both MDCT and QMF, the available time-frequency behavior is a compromise between a match to sustained stationary inputs and the ability to handle true input discontinuities in time or frequency.

Fig. 11(c) is a schematic of a perceptual audio coder based on subband or transform coding. The block diagram of Fig. 11(c) includes the possibility of jointly coding the two channels in a stereo-pair to maximize the gains due to redundancy removal and perceptual tuning [6], [66], [137].

C. DCT and Motion-Compensated DCT Coding of Image and Video

As in the case of audio, the absence of a strong autoregressive (LPC) model for the source signal has led to the well-accepted use of subband and transform coding methods for image and video compression. Source redundancy is addressed by decomposing the input signal into components of differing variance in the frequency or transform domain and following this by variable bit allocation in the quantization of the transform coefficients. A greater number of bits is allocated to the components of higher variance, and the overall mean squared error is minimized for a given constraint on total bit rate. Perceptual matching can be realized, at least partly, in the bit allocation process if a good model is used for the best possible profile of distortion versus frequency.

Of particular interest to image processing and image coding standards is the two-dimensional discrete cosine transform (2-D DCT) [1], [62], [100], [147]. The DCT provides a good match to the optimum (covariance-diagonalizing or Karhunen–Loeve) transform [62] for most image signals, and fast algorithms exist for computing the DCT.

A ubiquitously used frequency-selection or bit-allocation rule for 2-D DCT is one based on a lowpass process described by the *zig-zag* scan of Fig. 12(a), which refers to an 8 × 8 DCT operation with 64 samples of 2-D frequency ranging from DC (0, 0) to the highest frequency (7, 7). The low frequencies in the initial parts of the scan are given higher priority for retention or bit allocation.

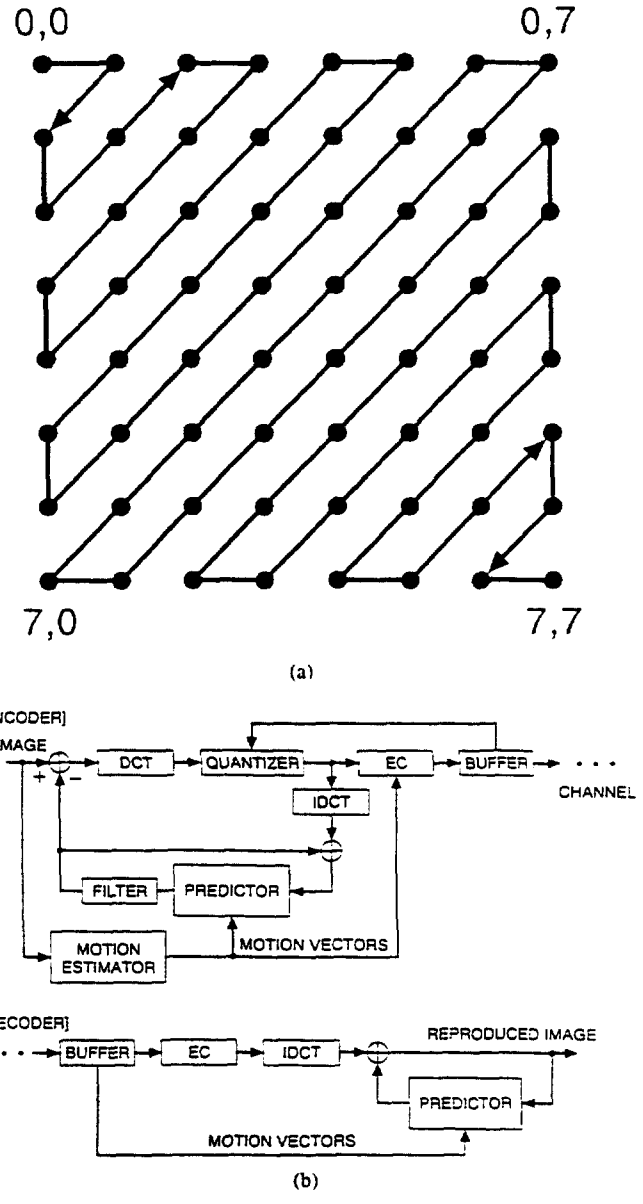


Fig. 12. (a) The zig-zag scan for frequency selection and bit allocation processes in DCT coding. (b) Block diagram of video coding based on motion compensation followed by DCT coding of the interframe residual (IDCT: Inverse DCT. EC: Entropy Coding) (after [81], [85], [100]).

and the zig-zag scan represents a monotonic deemphasis of the higher 2-D frequencies. More sophisticated algorithms (Section IV-B) can provide for a nonmonotonic bit allocation process for those inputs that demand it.

Fig. 12(b) is a block diagram of a video coding system based on interframe motion compensation followed by 2D-DCT coding of the residual signal after motion compensation. Motion vectors are derived from the input process and are used to define a space-varying interframe predictor. Abrupt changes in hypothesized motion are prevented by means of a filter before computing the motion-compensated interframe residual that is quantized by the 2-D DCT module. The transmitted information consists of the quantized residual and the motion vector. An entropy coder (EC) is used to reduce the redundancies in the DCT and motion vector sequences. Important sources of these redundancies are the unequal probability distributions in the possible ranges of quantized DCT components and the quantized motion vectors.

The system of Fig. 12(b) is widely used in video coding. It forms the basis of two international coding standards: the $p * 64$ kb/s standard of the CCITT ($p = 1$ to 24) [85] and the 1.1 Mb/s standard of ISO-MPEG [81]. The CCITT standard is intended for videotelephony and videoconferencing applications. The MPEG standard is intended for addressable video in multimedia applications. As such, the MPEG standard also allows for high-quality *intraframe* coding of selected frames in the video sequence. The MPEG algorithm also permits greater processing delays than the CCITT system, given the storage focus in MPEG.

Removal of the motion compensation loop in Fig. 12(b) results in a still-image coder based on 2-D DCT. The resulting system forms the basis of the international coding standard known as ISO-JPEG [147].

D. The Next Generation of Signal Coders

The techniques of Sections III A–C are low bit-rate systems with several important features. With the exception of the LPC vocoder, these systems are phase-preserving waveform coders with an emphasis on naturalness in output signal quality. These coders use generic techniques for reducing redundancy and for matching the quantizing system to the human perceptual mechanism. They generally include means for variable bit rate coding, provide for some degree of inherent resistance to transmission errors, and permit efficient implementation in existing signal processing technology. They form the bases for various international standards.

As we look toward the next generation of coding algorithms, we seek to decrease the bit rate even further for specified levels of signal quality and, in some applications, we need to increase reproduced signal quality at a specified bit rate. We need to develop the possibility of increasing the signal bandwidth that can be realized at a given bit rate and at a given level of quantizing distortion. Finally, we need to drive several technologies towards the ideal of (perceptually) distortion-free coding, especially speech and video technologies that currently provide only *communications quality* rather than *high* or *transparent* quality.

IV. Research Directions

In discussing directions of research, it is impossible to be exhaustive; and in predicting what the successful directions may be, we do not necessarily expect to be accurate. Nevertheless, it may be useful to set down some broad research directions, with a range that covers the obvious as well as the speculative. The remaining parts of Section IV are addressed to this task.

A. Subsampling, Interpolation, and Multiresolution Processing

Signal distortion in digital coding is a combination of prefiltering, aliasing, and quantizing components. The tradeoffs among these components are not rigorously understood. Typically, at high and moderate bit rates, prefiltering distortion is effectively a nonissue in that users accept or get used to a predefined and well-accepted bandwidth, as in telephone-grade speech. Aliasing distortion can be very noticeable even when it is small in a mean squared error sense. Typically, aliasing components are either avoided by proper prefiltering or, in the case of relatively complex arrangements such as subband coding, these components are carefully minimized. Quantizing distortion is, therefore, the typically most relevant part of reconstruction error; it is the component that one typically seeks to minimize in designing a coding algorithm at a specified bit rate.

As we consider lower bit rates and higher quality, all the components of coding distortion can become significant, and issues of their interaction and of relative tradeoffs become important as well.

One example of such a tradeoff is that between bandwidth (or sampling rate) and the bits per sample for a given total bit rate. With the possible exception of speech coding, where the telephone bandwidth of 3.2 kHz can be regarded as an essential minimum, the notion of optimizing bandwidth for a given bit rate can be quite an interesting problem. A good example is the definition of optimum bandwidth in audio given an overall low bit rate such as 16 or 32 kb/s. Another example is the definition of optimum spatio-temporal resolution in the coding of video signals at low bit rates such as 64 or 128 kb/s. Specific low resolutions such as CIF and quarter-CIF are selected using reasonably good criteria in low bit rate coding, and displayed signal resolution is enhanced by means of interpolation especially in the time domain [81], [85]. However, fundamental unanswered problems remain, such as the selection of the best fixed set of spatial and temporal resolutions for a given bit rate and the definition of efficient multiresolution algorithms [7], [28], [88], [141] for dynamic adaptation of spatial and/or spatio-temporal frequency resolutions. Rigorous solutions of these problems are difficult even with the simplifying assumption of zero aliasing, but refined models of coding and psychophysics may be able to translate what is currently no more than an art to a science with a reasonable potential for generalization.

Although there is little flexibility in the bandwidth or sampling rate of telephone speech, as mentioned earlier, the techniques of subsampling and interpolation are still extremely valuable in the spectral domain. For example, the interpolation of LPC parameters is a widely practiced tool for low bit-rate speech coding [41], [71], [79]. More recently, an interpolation technique for the compact description of the excitation waveform of voiced speech has been proposed. The technique, called *prototype waveform interpolation* [76], is based on the transmission of a prototype excitation waveform and its pitch once every update time, on the order of 20 to 30 ms. A complete excitation signal is obtained by means of interpolative techniques in the Fourier series domain. The process preserves a high level of periodicity, and is flexible enough to realize low levels of reverberation and tonal artifacts. Combined with CELP for the coding of unvoiced speech, the technique provides a promising framework for high-quality coding at rates on the order of 2 to 4 kb/s.

B. Techniques for Time-Frequency Analysis

The assumption of zero aliasing is a conspicuous simplification in almost all subband coding literature. Techniques such as quadrature-mirror filtering and the modified discrete cosine transform can provide perfect cancelation of aliasing in the absence of quantization errors. This ideal solution never occurs in a practical coding situation, and the assumption of the ideal case becomes increasingly inappropriate at lower bit rates because of a corresponding increase of quantization error. Recent work has given us a fairly good understanding of filterbanks that provide zero or near-zero reconstruction error in the absence of quantization. However, the design of a filterbank that minimizes the combined effect of quantizing and aliasing errors is an entirely unsolved problem. Here again, a rigorous optimization is extremely intractable in our current state of knowledge but we sorely need at least partial solutions.

A somewhat orthogonal, though not unrelated, research area is that of efficient time-frequency analysis. Considerations of signal nonstationarity and perceptual distortion criteria have resulted in increasingly sophisticated frameworks for signal analysis. In particular, techniques that provide flexible combinations of time-support and bandwidth represent a powerful generic tool for efficient coding. The discrete Fourier transform and a uniform-bandwidth quadrature mirror filterbank are well-understood and widely used analysis tools. But in their simplest forms, they lack the flexibility for time-frequency analysis mentioned above. QMF trees with unequal-bandwidth branches, as well as subband-DFT hybrids, are relatively newer structures with more flexible features. This is also true of *wavelet* filters [21], [88], [120].

Wavelets: Unlike the basis vectors of a DFT (sinusoids and cosinusoids of various frequencies and constant time-support), the wavelet filter structure is characterized by a shorter time support at higher frequencies and a longer time support at lower frequencies, a direct result of a dilating operation that is a basic component of wavelet design (Fig. 13). The time-frequency characteristic of a wavelet filterbank is a natural match to some of the properties of audio-visual information: high-frequency events often occur for a short time and stand to benefit from a finer resolution in time analysis, while low-frequency events are often sustained in time and require less frequent sampling in time. The wavelet approach, especially if used in a time-varying framework, may therefore offer

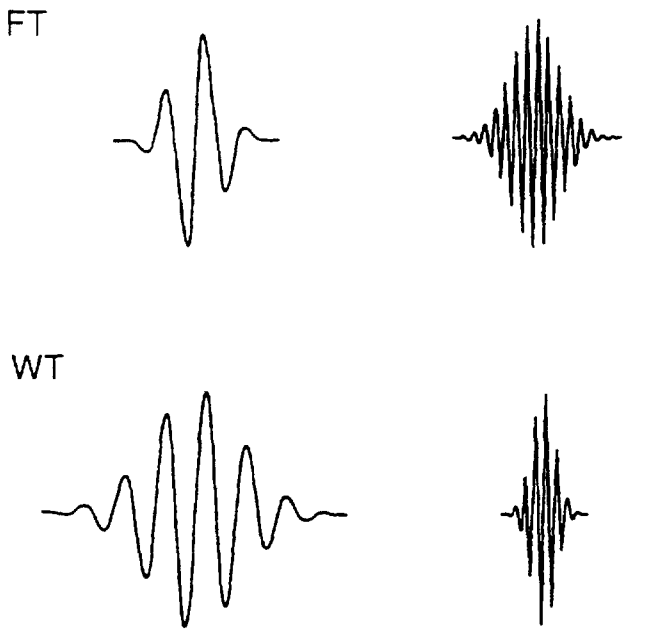


Fig. 13. Qualitative comparison of Fourier and wavelet transforms (after [120]).

powerful forms of adaptive analysis in a more basic sense than a nonuniform frequency-band QMF system or a variable window-length MDCT system. Wavelet transforms provide the additional feature of perfect reconstruction, a property not generally offered in conventional methods of analysis. Conventional methods, however, do offer the property of *almost-perfect* reconstruction which is adequate for many low bit rate applications.

Wavelet filtering is a promising analytical tool and applications of it are also beginning to emerge. What is still lacking, however, is a thorough understanding of what wavelets can do for coding that the more sophisticated examples of conventional analyses cannot. As we seek to apply wavelet (or nonwavelet) tools to low bit rate coding, attention must necessarily shift to the yet-untouched problems of uncanceled aliasing and the computationally-intensive but naturally appealing notion of a signal-adaptive filterbank.

Multidimensional Subband Coding: The discussion in Section III-C implied a two-dimensional DCT (2-D DCT) for still-image coding and the hybrid approach of motion compensation combined with 2-D DCT for the coding of video. Alternative approaches in current research include 2-D subband coding for still-image coding [25], [124], [150], [154] and 3-D subband coding for video [69], [112], [144]. The subband filters in each case may belong to any desired class, such as QMF's or wavelets. As in our discussion of (1-D) QMF and (1-D) MDCT techniques for audio, we note that (2-D and 3-D) transform and subband techniques are, in principle, dual operations. However, in the context of an overall coding system, differences can exist in terms of implementation, matching to the human perceptual system and robustness to transmission errors.

Fig. 14 shows the analysis filterbanks for 2-D and 3-D subband coding using separable (but not necessarily identical) filters in the different dimensions. In the illustrated examples, the 2-D filterbank provides a 16-band partition of the spatial 2-D frequency space and the 3-D filterbank provides an 8-subband partitioning of the spatio-temporal 3-D frequency space. We shall comment again on the 2-D subband coder in the context of perceptual coding (Section IV-D-2). The 3-D subband coder represents a significant departure from the current practice of motion compensation followed by a spatial transform. Rather than comparing two adjacent frames to realize good models of local motion, the 3-D approach processes two or more adjacent frames in order to capture spatial and temporal activity in a more integrated fashion. Most of the signal energy tends to be in the subband with low temporal and horizontal frequencies. Motion detection occurs prominently in the subband with high temporal frequency and low spatial frequencies. Low bit rate coding results from a variable bit allocation algorithm that matches not only the energies in the spatio-temporal subbands but also the respective measures of perceptual significance (ideally, a spatial-temporal 3-D JND profile).

C. Vector Quantization

It is generally recognized that in the compression of audio and visual signals, a suitably global distortion metric is more meaningful than a local single-sample-oriented metric. Simple forms of waveform coding, such as PCM and differential PCM (DPCM), are based on a local distortion metric. More complex coders, such as delayed-decision DPCM and block-oriented coders of various kinds, are characterized by the use of a more global distortion criterion. Examples of block coders are CELP systems for speech coding and 1-D and 2-D transform coders for audio and image signals. The general mechanism that permits the use of a global distortion criterion is called *delayed decision coding* [20], [62].

Vector quantization (VQ), tree coding, and trellis coding are all techniques for delayed-decision coding [62]. Unlike the block-oriented approach in VQ, tree and trellis structures use a sequential procedure to minimize a suitably global distortion measure. The trellis structure, which can be realized by using a finite-impulse-response code generator, has the advantage that the total number of possible output values is a finite number by definition, permitting efficient searches for the best path through the trellis (the coded output sequence) [33]. The VQ paradigm also has a finite output alphabet, by definition.

Several examples of vector, tree, and trellis coders have appeared in coding literature [62], together with some examples of hybrid algorithms using both block and sequential approaches [107]. The block and sequential approaches are not fundamentally different from a rate-distortion viewpoint. The block approach of VQ seems, however, to be somewhat better understood and more widely practiced, with a broad repertoire of special techniques for codebook design, fast codevector search, in-

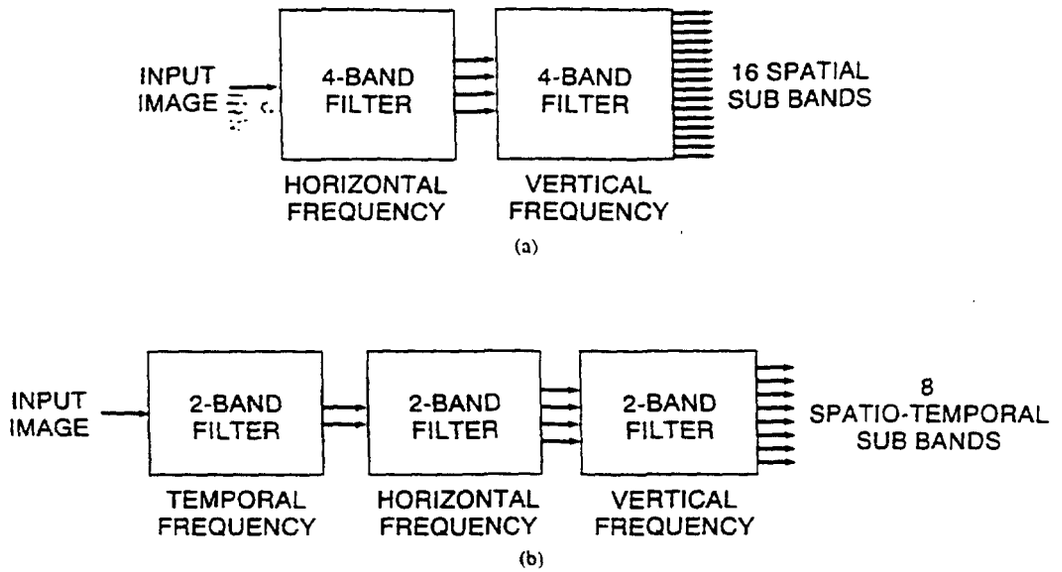


Fig. 14. Examples of analysis filterbanks for: (a) two-dimensional and (b) three-dimensional subband coding (after [124] and [69], [112]).

terblock coding, and optimization for noisy transmission channels [11], [14], [24], [27], [28], [38], [39], [40], [48], [49], [51], [72], [82], [83], [84], [92], [105], [107], [112].

Vector quantization (VQ) is known to be most efficient at very low bit rates (on the order of $R = 0.5$ bit per sample and less). This is because when R is small, one can afford to use a large vector dimension N and yet have a reasonable size 2^{RN} of the VQ codebook. Use of a large dimension N tends to bring out the inherent capability of VQ to address linear as well as nonlinear redundancies [87] in the components of the vector being quantized.

In speech coding, the use of VQ has been most successful in the coding of parameters describing the speech spectrum (LPC-VQ) [37], [67], [87], [109], [123] and speech excitation [4], [12], [13], [41]. Refinements such as gain-shape VQ [125] and multistage VQ [67], [83] have been routinely applied in the quantization of the spectrum and excitation parameters in speech. Direct VQ coding of the waveform has been less successful to date. This is due partly to intervector discontinuities resulting from a block-quantization process, and partly to the lack of a perceptually good model for minimizing intravector distortion. With LPC-VQ and CELP-coding, the distortion in the VQ process is conveniently frequency-shaped. This is accomplished by a spectral distortion metric in LPC quantization, and by the subsequent LPC filter itself in the case of excitation VQ. Direct VQ coding of the speech waveform, while not very successful to date, is a very interesting research problem and it offers an excellent challenge to the notion of adaptive and perceptually-tuned vector quantization.

Scalar quantization, followed by entropy coding [54], [55], [81], [85], [153], has some equivalences with vector quantization, at least in terms of attaining high values of signal-to-noise ratio at relatively low bit rates, especially if the probability density function of the input signal is Laplacian or gamma (rather than, say, uniform or Gaussian). This equivalence has perhaps diminished the application of VQ. However, powerful algorithms result when selected vectors from a DCT process are identified for low bit-rate vector quantization. Such techniques for DCT-VQ coding have been used successfully in image and audio coding [11].

Direct VQ coding of the signal waveform has been relatively more successful in the coding of intensities in 2-D images [49], [51], [96], [99], [118], [125], [141], [149], [150], [155] and in the coding of interframe 2-D residuals in motion-compensated coding of video. In the usual case of iteratively trained codebooks, the need for a typical database for training and the possibility of a significant mismatch between the input sequence and the trained VQ codebook are both important, if not problematic, issues. The importance of finite-state vector quantization to provide 2-D adaptivity of the codebook has been well demonstrated, resulting for instance in high-quality coding of a still image at 0.5 bit per sample. In the case of high-frequency image subbands dominated by sparse intensity profiles such as edges, an untrained system called *geometric vector quantization* [112] has been shown to provide very efficient coding at extremely low bit rates, on the order of 0.1 bit per sample and less.

As we move toward the next generation of low bit-rate algorithms for image and video coding, more ubiquitous use of vector quantization is very likely. Further research is needed to support these applications. We need a better understanding of tradeoffs and interactions between vector quantization and entropy coding, better experience with VQ-type structures such as pruned trees, better algorithms for adaptive VQ in 1, 2, and 3 dimensions, and more powerful perceptual models for characterizing and minimizing the vector-distortion process.

In expanding the frontiers of VQ research, a potential source of cross-fertilization is the field of *fractal block coding* [57]. Recent work on the subject has shown image quality similar to that of 2-D VQ in the coding of still images at rates on the order of 0.5 bit per sample. The fractal method utilizes a subtle form of self-similarity in a scene, in particular, similarities between selected pairs of blocks in 2-D images in the presence of a powerful set of transformations. In each such pair, the block to be encoded is called a child and the potential matching reference is called a parent (Fig. 15). Allowed transformations include changes of scale, orientation and grey level, or color. An image is encoded by means of a fractal code that consists of a sequence of child–parent maps and a corresponding sequence of transformations chosen from a transformation codebook. The child–parent map and transformation, for any given input block, are results of a joint optimization that minimizes the error in child–parent matching. Since there is no insistence on a zero-matching error (perfect child–parent similarity), the method can be called a *soft-fractal* technique. Decoding consists of iterative calls of the fractal code on an arbitrary initial image. The block-fractal algorithm can be viewed as an interblock coding algorithm with nonlinear block prediction. In this sense, it is reminiscent of predictive and finite-state vector quantization algorithms [48], [51], [72], [107]. It appears that new research on adaptive quantization may enrich the fields of fractal coding and VQ alike, and that these disciplines, as we know them today, may also have mutually orthogonal strengths.

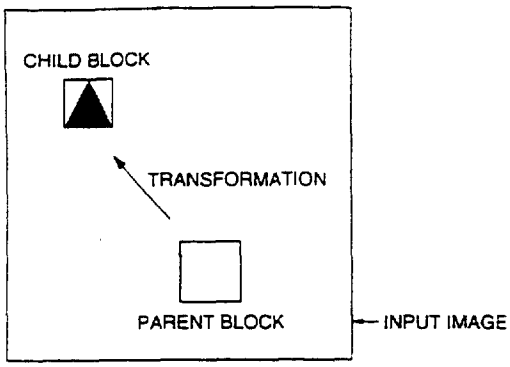


Fig. 15. Image coding based on soft-fractal analysis (after [57]).

D. Models of Signal Production and Perception

Central to the success of the techniques discussed in Sections IV A–C are models of signal production and perception. These models need to be physically realistic and computationally feasible. In some problems, as in filter-bank design, it is desirable to seek architectures that are matched reasonably both to the production and perception models. On the other hand, it is important to realize that entirely different requirements may exist in the matching processes mentioned above. In order to address the complex and not identical phenomenon in signal production and signal perception, it is important that the signal processing architectures used in signal coding are as general and flexible as possible.

1) Models of Signal Production: As mentioned earlier, the speech waveform is unique in that it enjoys a reasonably well-understood and universal source model that is efficiently approximated as the result of an excitation signal modulated by a linear transfer function. In the search for high-quality LPC coders, recent algorithms such as CELP have gotten away from the restrictive voiced-unvoiced binary model for excitation. However, as we increase the focus on very low bit rates such as 2–4 kb/s, the voiced-unvoiced model is receiving renewed attention because of its compactness and the resulting economies in bit rate. These include the techniques of multiband excitation coding [50], [52], sinusoidal coding [90], and prototype-waveform interpolation [76]. In the new generation of coders based on the voiced-unvoiced classification, the attempt is to make the classification soft and robust, so that the effect of a wrong classification is imperceptible. A generalization of the binary voiced-unvoiced classification is a system where classification is based on a larger number of states and on a phonetic criterion for state-switching [148].

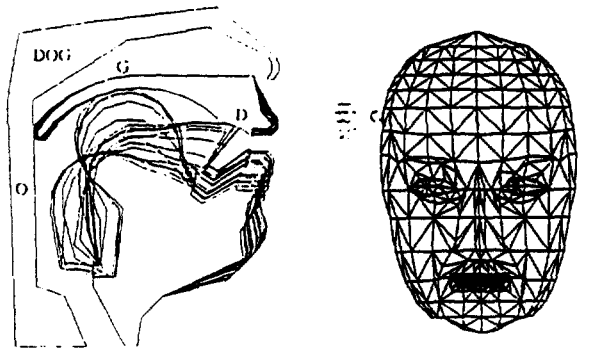
Indirect analogies to the voiced-unvoiced model exist in audio and image coding as well. In audio, the classification of an input signal as a tone-like or a noise-like signal provides a good basis for adaptive perceptual coding. More powerful adaptation results if the signal can be associated with degrees of tonality and noisiness along a continuum.

In image processing, an analogous classification is into categories dominated by edges, textures, and flat regions of grey [118]. Another kind of classification results from segmenting an image into background and moving areas. a classification that is particularly meaningful in a typical teleconferencing scene. Resulting image classes call for different algorithms, redundancy removal, and perceptually-matched quantization.

As we seek to extend the capabilities of the above models, especially the less mature models of audio and image, we need to optimize prediction, transform, and quantization processes for the various signal classes. We also need to address the problem of maintaining perceptual continuity in a signal in the context of the breaking up of the signal into several sets of homogeneous parts in model-based classification processes [68], [80], [111].

In principle, the ultimate results in coding are those obtained when the models reflect the very earliest stages in the signal production processes. Examples are the articulatory vocal cord–vocal tract model of human speech production and the wire-frame image model of a human face (Fig. 16).

The articulatory model [128] of Fig. 16(a) extends the focus from LPC analysis to the analysis of vocal tract areas and, in principle, provides a much stronger domain for very low bit-rate vector quantization. The model also permits a better handle on the interaction between vocal cords and the vocal tract, a phenomenon that is conspic-



(a) (b)

Fig. 16. (a) Articulatory model for speech coding (after [128]). (b) Wire-frame model for coding images of the human face (after [2], [94]).

uously ignored in the simplified traditional model of excitation-followed-by-filter.

The wire-frame model of the human face [2], [29], [94] in Fig. 16(b) is potentially a very powerful domain for compressing scenes dominated by the face image, as in a videotelephone scene with a closeup of the human face. The general wire-frame model is a 3-D lattice with a very large number of polygonal sections. Using affine transformations and the actual image of a given face, a personal facial model is created. Vector-quantized versions of dynamically changing mouth features driven by textual or speech cues, together with algorithms for representing text- or speech-independent facial features, are used to create a talking-head image.

The above models are a natural basis for an intelligent human–machine interface. If they are intended as a basis for human telecommunication, an important obvious challenge for articulatory and wire-frame models is the realization of natural, rather than synthetic, signal quality at the very low bit rates such models are intended for.

2) Models of Signal Perception: Perceptual criteria have been addressed since the very beginnings of speech and image coding, and as coding algorithms have matured, criteria for optimizing these algorithms have become increasingly complex [3], [6], [30], [58], [66], [98], [102]–[104],. [110], [121], [124], [126], [127], [135], [137]. An interesting early example of perceptual coding is the use of dithering to improve the quality of low bit-rate PCM by breaking up structured and, hence, highly visible patterns in the distortion process [62], [121]. Examples of complex perceptual coders are the time- and space-varying algorithms discussed in the remainder of this section.

Models of human hearing and vision can steer a coding algorithm in two related but distinct ways. If one can define, for each part of a signal being coded, a just-noticeable-distortion (JND) level below which reconstruction errors are rendered imperceptible because of masking, the model will be the basis of perceptually lossless coding, a process in which the perceived distortion D_p is zero even if a mathematically measurable distortion such as the mean-squared-error D_{mse} is nonzero (in fact, quite significantly so in typical examples). If, on the other hand, the average bit rate for coding is not sufficient to realize the JND profile in all parts of the signal, the perceptual model can still suggest a good match of the quantizing system to the perceptual model, in the sense of minimizing D_p, rather than D_{mse}. In this case, the input to the quantizing system is not a JND profile as in Fig. 11(a) but a *minimally-noticeable-distortion* or MND profile, as in Fig. 10.

A generic block diagram of a perceptual coder driven by JND (or MND) cues is shown in Fig. 17. The JND and MND profiles are meant to be dynamic, being functions of local signal properties such as dominant frequency, background intensity or texture, and local temporal activity. The mapping from these properties to the JND or MND profile is performed in real time, although the function defining the mapping is established prior to coding, based typically on extensive subjective experimentation. In the case of the JND coder, the system is a constant-quality, variable-bit-rate system by definition. However, feedback from a bit rate buffer can be used to realize a constant-bit-rate variable-quality system whose distortion profile approximates the JND. In a conservative design, the actual distortion would be less than the JND most of the time and greater than the JND on occasion.

The JND method has been extremely successful in the transparent and near-transparent coding of wideband audio [Fig. 11(a)] and, more recently, in the transparent coding of still images based on 2-D subband analysis at extremely low bit rates. Fig. 18 illustrates the JND as a function of space for the example of a subband image signal. Parts (a)–(c) of the figure display: (a) the input image; (b) the lowest frequency subband of it (low horizontal and low vertical frequencies in the 16 subband system of Fig. 14(a); and (c) the JND image for the coding of that subband. Following the coding of the 16 subbands based on respective JND profiles, the coded subbands are combined in the synthesis filter to obtain a low bit-rate coded image. Part (d) of the figure describes, as a function of space, the number of subbands (out of 16) that are retained (with nonzero bit allocation) in the JND-driven system. White, light grey, dark grey, and black blocks retain 4, 3, 2, and 1 subbands out of 16, respectively, indicating a high degree of compression.

The perceptual subband coder results in excellent image quality at extremely low bit rates, generally lower than the rates realized by the well-known ISO-JPEG DCT algorithm. This capability has also led to a new proposal for the coding of high-resolution facsimile—as a grey-level input compressed to extremely low rates and half-toned at the receiver rather than at the transmitter [104]. Retention of the grey-level domain in most of the system has been shown to result not only in lower transmission time on a given digital channel but also in a significantly sharper end-result with a printer of given resolution, provided that the half-toning algorithm is optimized using models of the printer and the human visual system [110].

In the coding of 3-D images, a major research challenge is the definition of temporal models of masking and

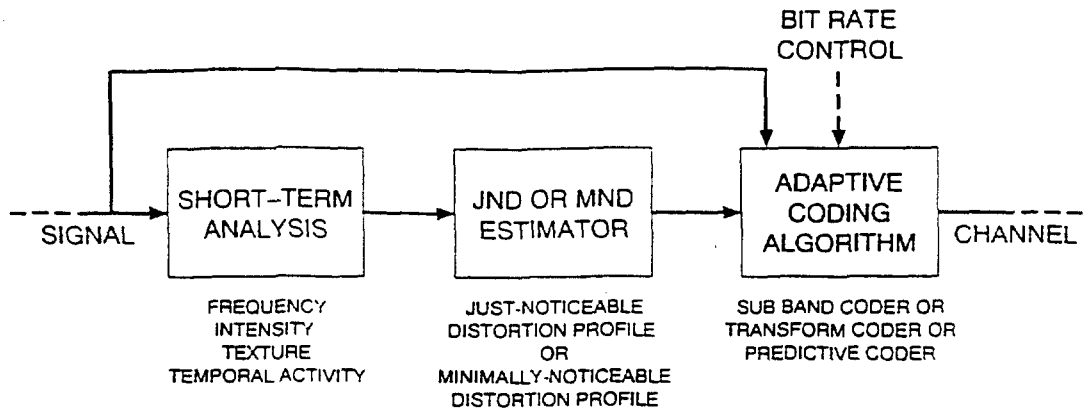


Fig. 17. Generic block diagram of perceptual coding.

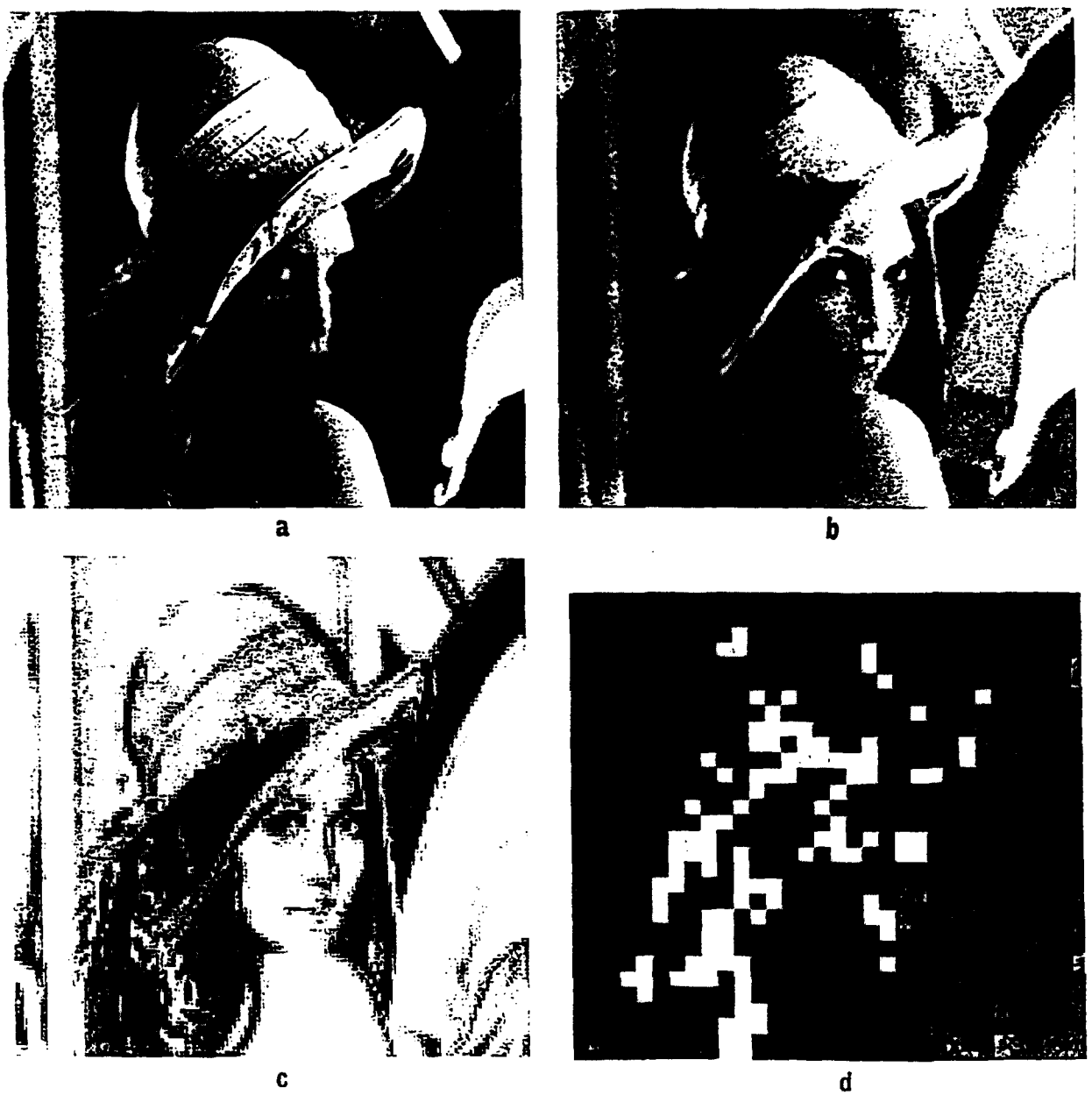


Fig. 18. Perceptual subband coding. (a) 512 × 512 image of *Lena*. (b) Low-frequency subband of *Lena* in a 16-band 2-D subband coder. (c) The JND image for (b). (d) The spatial bit-allocation profile (after [124]).

the definition of a suitable coding framework for exploiting these physical models. Recent experiments in motion-compensated coding for high-definition television have shown the potential of spatial masking models as in Fig. 18, as well as temporal masking models that describe the masking of distortion by certain kinds of motion activity [103]. It is possible that 3-D subband analysis [Fig. 14(b)] will provide an even better description of the motion process and, therefore, a better framework for perceptual coding [112].

Temporal models of noise masking are also of great interest in the coding of audio and speech. More powerful masking models are crucial to the goal of high-quality coding of speech at very low bit rates. The focus here is on MND functions that provide increasingly more efficient formulas for noise shaping. JND formulas resulting in transparent coding (at slightly higher bit rates) are expected to be byproducts rather than prime targets in low bit-rate telephony.

An MND-based coder has an easier function, in principle, than a JND-based coder because a nonzero value of the perceptual distortion D_p is allowed. However, the methodology leading to a good MND design can actually be more complex than that leading to a JND design at a higher bit rate. This has to do with the subjective experiments that are the bases for the JND and MND formulas. In such experiments, it is easier to identify a situation where D_p is zero (unnoticed distortion by a specified percentage, say 50 or 95, of the subject population) than to associate a perceptually meaningful value, along a continuum, to the distortion (given that it is clearly noticeable). For the same reason, if one were to seek a distortion-rate description of the signal, as in information theory, it is easier to identify the bit rate $R_p(0)$ required for $D_p = 0$ than to calculate a perceptually valid shape for the entire $D_p(R)$ curve. We expect, of course, that this curve is a significant left-shift of the traditional curve of D_{mse} versus R (Fig. 19).

The mean squared distortion at zero bit rate is equal to the signal power itself. The bit rate $R_{mse}(0)$ at which the mean squared error is zero is infinite for a continuous-amplitude source, and finite for a discrete-amplitude source such as a computer-stored file of 8-bit image pixels ($R_{mse}(0) = 8$ in this case; in fact less, typically in the neighborhood of five bits per sample if mathematically lossless coding is performed). The bit rate $R_p(0)$ at which the perceived distortion can be designed to be zero is a fundamental limit in coding, and we will call it the *perceptual entropy* of the signal. More generally, the $D_p(R)$ curve defines the fundamental limit of signal compression for a specified level of output signal quality.

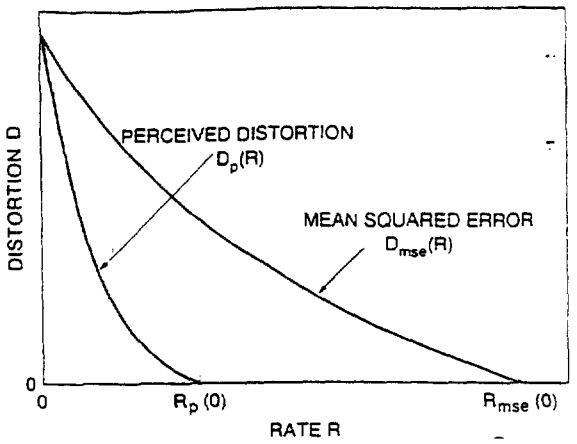


Fig. 19. Distortion-rate function for mean squared error and perceived distortion.

E. Source and Channel Coding

Traditionally, source and channel coding have had complementary roles in digital communication. The source coder has tried to minimize the bits-per-sample for high-quality signal representation, while the channel coder (modulator and error protection system) has attempted to maximize the bits-per-second-per Hertz that can be used on a transmission or storage medium to represent the digitally coded signal. While these complementary roles will continue in future technology, it will become increasingly important to define interactive and joint designs of the source and channel coding algorithms.

For example, if the output of a source coder can be categorized into parts of varying sensitivity to bit errors in transmission, a given total overhead for error protection can be used in an unequal error protection scheme that will have the final effect of extending the range of channel quality over which a specified quality of signal communication can be maintained [Fig. 20(a)] [17]. In situations where the transmitter has information about channel quality, a joint source-and-channel coding algorithm can make a suitable allocation of the total bit rate for source coding and error protection [93]. This, again, has the effect of utilizing transmission media at low levels of channel quality: a slight undercoding of the signal in a quantization-noise sense, together with a stronger focus on error protection, can realize a specified level of total (quantization-plus-channel) noise over a wider range of channel quality [Fig. 20(b)]. Finally, by refining our methods for source coding, channel coding, and cooperative source-and-channel coding, we can generally enhance the gains over analog communication [Fig. 20(c)]. Recent examples of this appear in digital cellular technology for mobile radio telephony and digital transmission proposals for high-definition television. Ideas of perceptual optimization, discussed in earlier sections of this paper for signal coding, carry over to signal communication as well, as in optimizing algorithms for unequal error protection and joint source and channel coding.

F. Signal Compression and Packet Networks

Earlier work on packet speech and video was concerned with the effects of packet losses and means for interpolating the signal in the presence of packet losses [45], [61], [70], [119], [142]. This line of research is being reacti-

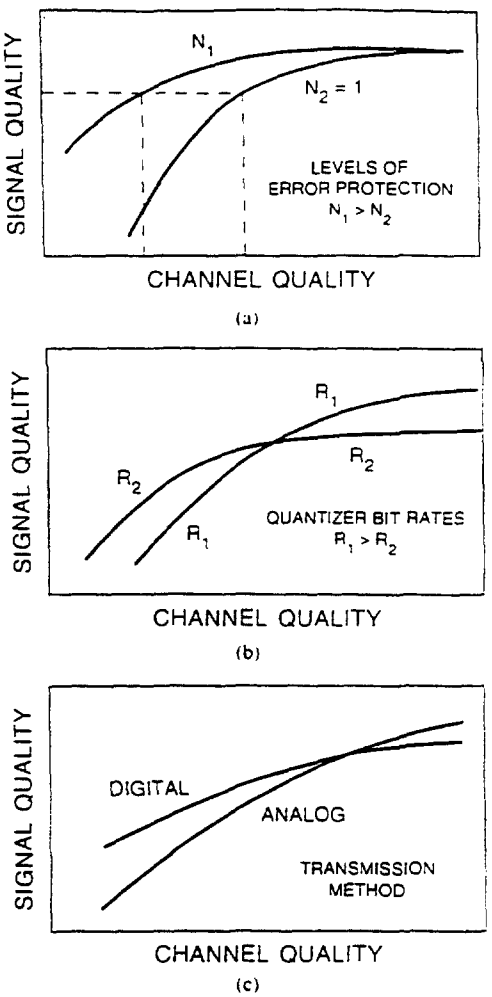


Fig. 20. Signal quality as a function of transmission channel quality, with: (a) digital transmission using unequal error protection; (b) digital transmission using joint source and channel coding; and (c) analog and digital transmission.

vated today because of the increasing importance of the asynchronous transfer mode (ATM) packet technology. An important concept here is that of *layered coding*. The output of the source encoder is divided into cells of varying significance, typically with two layers, or levels of it. When the packet network is congested, the idea is to drop cells of lower priority [75], [95], [138], [143]. In the case of uncompressed PCM data, the prioritization of encoder bits is straightforward. But in low bit rate coders, as in the unequal error protection scheme of Fig. 20(a), optimization of the communication network will depend on perceptual cues for cell layering and subjective methods for measuring the user acceptance of layered coding.

V. Conclusion

This paper has presented a description of technology targets in signal compression and a nonexhaustive, and very possibly biased, account of research directions that may lead us toward these targets. As we pursue these and other directions, some of which are undescribed here and some of which are quite unknown at this time, one broad observation would perhaps stay uncontested: the new generation of algorithms will reflect, better than ever before, perceptual cues as integral parts of the coding process. In order to calibrate and steer our research progress in a discipline where the performance criteria are mathematically hard to model, we will also depend increasingly on subjective evaluations of signal quality, as in Fig. 9.

Quality evaluations have been invaluable in speech and audio coding and in the early history of television, but conspicuously lacking in contemporary work on digital image coding. Experiments to define specific distortion models for optimizing a coder, and experiments to measure the overall quality of the coded signal, are both very time-consuming and intricate. However, both of these endeavors will be necessary investments if we seek to advance signal coding technology to the fundamental limits defined by information theory and psychophysics.

This paper has also pointed out opportunities for integrating source coding and channel coding technologies. Such integration, which has hitherto been an informal exercise, will become increasingly essential as we stretch our communication capabilities with capacity-limited channels such as wireless media. In parallel, as we seek greater sophistication in the integration of speech and data with broadband signals such as CD-audio and high-resolution video, we will witness an increased interaction of signal compression technology with the field of communication networking.

Acknowledgment

The author is indebted to three anonymous reviewers whose extensive and constructive criticism has added significantly to the value and credibility of this article. Also, for their invaluable inputs, many thanks to his esteemed colleagues P. Noll, L. Rabiner, B. Atal, J. Johnston, N. Seshadri, S. Quackenbush, and J. Schroeter. Thanks also to T. Pappas for providing the pictures in Fig. 18.

References

[1] N. Ahmed, T. Natarajan, and K. Rao, "Discrete cosine transform," *IEEE Trans. Comput.*, pp. 90–93, Jan. 1974.

[2] K. Aizawa, H. Harashima, and T. Saito, "Model-based analysis-synthesis image coding (MBASIC) system for a person's face," *Signal Processing: Image Communication*, New York: Elsevier Science, Oct. 1989, pp. 139–152.

[3] B. S. Atal and M. R. Schroeder, "Predictive coding of speech signals and subjective error criteria," *IEEE Trans. Acoust., Speech and Signal Proces.*, pp. 247–254, June 1979.

[4] B. S. Atal, "High-quality speech at low bit rates: Multi-pulse and stochastically excited linear predictive coders," in *Proc. ICASSP*, 1986, pp. 1681–1684.

[5] T. Berger, *Rate Distortion Theory*, Englewood Cliffs, NJ: Prentice Hall, 1971.

[6] K. H. Brandenburg, "OCF—A new coding algorithm for high quality sound signals," in *Proc. ICASSP*, Apr. 1987, pp. 141–144.

[7] P. J. Burt and E. H. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Trans. Commun.*, vol. COM-31, pp. 532–540, Apr. 1983.

[8] CCIR, "Method for the subjective assessment of the quality of television pictures," Rec. 500-1, 1978.

[9] CCITT Study Group XVIII, "32 kb/s adaptive differential pulse code modulation (ADPCM)," Working Party 8, Draft Revision of

Recommendation G.721, Temporary Document No. D.723/XVIII, Source: USA: Geneva, Switzerland.
[10] CCITT Study Group XVIII, "7 kHz audio coding within 64 kb/s," CCITT Draft Recommendation G.722, Report of Working Party XVIII/8, July 1986.
[11] W.-Y. Chan and A. Gersho, "Constrained-storage quantization of multiple vector sources by codebook sharing," *IEEE Trans. Commun.*, vol. 39, pp. 11–13, Jan. 1991.
[12] J.-H. Chen and A. Gersho, "Real-time vector APC speech coding at 4800 b/s with adaptive postfiltering," in *Proc. ICASSP*, Apr. 1987, pp. 2185–2188.
[13] J.-H. Chen, R. V. Cox, Y-C. Lin, N. S. Jayant, and M. J. Melchner, "A low-delay CELP coder for the CCITT 16 kb/s speech coding standard," this issue, pp. 830–849.
[14] P. A. Chou, T. Lookabaugh, and R. M. Gray, "Entropy-constrained vector quantization," *IEEE Trans. Signal Proces.*, vol. 37, pp. 31–42, 1989.
[15] D. C. Cox, "Portable digital radio communications—An approach to tetherless access," *IEEE Commun. Mag.*, pp. 30–40, July 1989.
[16] R. V. Cox, "The design of uniformly and nonuniformly spaced pseudoquadrature mirror filters," *IEEE Trans. Signal Proces.*, vol. 34, pp. 1090–1096, 1986.
[17] R. V. Cox, J. Hagenauer, N. Seshadri, and C-E. W. Sundberg, "Subband speech coding and matched convolutional channel coding for mobile radio channels," *IEEE Trans. Signal Proces.*, vol. 39, no. 8, pp. 1717–1731, Aug. 1991.
[18] R. E. Crochiere, S. A. Webber, and J. L. Flanagan, "Digital coding of speech in subbands," *Bell Syst. Tech. J.*, pp. 1069–1085, 1976.
[19] C. C. Cutler, "Differential quantization for communication signals," U.S. Patent 2 605 361, July 29, 1952.
[20] ——, "Delayed encoding: Stabilizer for adaptive coders," *IEEE Trans. Commun.*, vol. 19, pp. 898–904, Dec. 1971.
[21] I. Daubechies, "Orthonormal bases on compactly supported wavelets," *Comm. Pure Appl. Math.*, pp. 909–996, 1988.
[22] W. R. Daumer, "Subjective evaluation of several efficient speech coders," *IEEE Trans. Commun.*, pp. 655–662, Apr. 1982.
[23] L. D. Davisson, "Rate distortion theory and application," *Proc. IEEE*, pp. 800–808, July 1972.
[24] J. R. B. deMarca and N. S. Jayant, "An algorithm for assigning binary indices to the codevectors of a multidimensional quantizer," in *Proc. ICC*, June 1987, pp. 1128–1132.
[25] C. Diab, R. Prost, and R. Goutte, "Block-adaptive subband coding of images," in *Proc. ICASSP*, Apr. 1990, pp. 2093–2096.
[26] D. Esteban and C. Galand, "Application of quadrature mirror filters to spit band voice coding schemes," in *Proc. ICASSP*, 1987, pp. 191–195.
[27] N. Farvardin, "A study of vector quantization for noisy channels," *IEEE Trans. Inform. Theory*, pp. 799–809, July 1990.
[28] T. R. Fischer, "A pyramid vector quantizer," *IEEE Trans. Inform. Theory*, pp. 568–583, July 1986.
[29] R. Forscheimer and T. Kronander, "Image coding—from waveforms to animation," *IEEE Trans. Signal Proces.*, pp. 2008–2023, Dec. 1989.
[30] J. L. Flanagan *et al.*, "Speech coding," *IEEE Trans. Commun.*, vol. 27, pp. 710–737, Apr. 1979.
[31] J. L. Flanagan, D. A. Berkley, G. Elko, J. E. West, and M. M. Sondhi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, 1991.
[32] J. L. Flanagan, *Speech Analysis, Synthesis and Perception*. New York: Springer-Verlag, 1972.
[33] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, pp. 268–278, Mar. 1973.
[34] A. Fuldseth, E. Harborg, F. T. Johansen, and J. E. Knudsen, "A real time implementable 7 kHz speech coder at 16 kbps," presented at Proc. Eurospeech 91, Genoa, 1991.
[35] S. Furui and M. M. Sondhi, Eds., *Advances in Speech Signal Processing*. New York: Marcell-Dekker, 1992.
[36] R. G. Gallager, *Information Theory and Reliable Communication*. New York: McGraw Hill, 1965.
[37] A. Gersho and V. Cuperman, "Vector quantization: A pattern matching technique for speech coding," *IEEE Commun. Mag.*, pp. 15–21, 1983.
[38] A. Gersho, "Asymptotically optimum block quantization," *IEEE Trans. Inform. Theory*, pp. 373–380, July 1979.
[39] A. Gersho, "On the structure of vector quantizers," *IEEE Trans. Inform. Theory*, pp. 157–166, Mar. 1982.
[40] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Kluwer Academic, 1992.
[41] I. A. Gerson and M. A. Jasiuk, "Vector sum excited linear prediction (VSELP)," presented at IEEE Workshop on Speech Cnding for Telecommun., Sept. 5–8, 1989.
[42] J. D. Gibson, "Sequentially adaptive backward prediction in ADPCM speech coders," *IEEE Trans. Commun.*, pp. 145–150, Jan. 1978.
[43] T. J. Goblick Jr. and J. L. Holsinger, "Analog source digitization: A comparison of theory and practice," *IEEE Trans. Inform. Theory*, pp. 323–326, 1967.
[44] R. C. Gonzalez and P. Wintz, *Digital Image Processing*. Reading, MA: Addison-Wesley, 1977.
[45] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W-C. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communication," *IEEE Trans. Signal Proces.*, pp. 1440–1448, Dec. 1986.
[46] D. J. Goodman, "Speech quality of the same speech transmission conditions in seven different countries," *IEEE Trans. Commun.*, pp. 642–654, Apr. 1982.
[47] D. J. Goodman, "Embedded DPCM for variable bit rate transmission," *IEEE Trans. Commun.*, pp. 1040–1066, July 1980.
[48] R. M. Gray, J. Foster, and M. O. Dunham, "Finite-state vector quantization for waveform coding," *IEEE Trans. Inform. Theory*, pp. 348–359, 1985.
[49] R. M. Gray, "Vector quantization," *IEEE ASSP Mag.*, pp. 4–29, 1984.
[50] D. W. Griffin and J. S. Lim, "Multiband excitation vocoder," *IEEE Trans. Signal Proces.*, pp. 1223–1235, 1988.
[51] H.-M. Hang and J. W. Woods, "Predictive vector quantization of images," *IEEE Trans. Commun.*, vol. 37, pp. 1208–1219, 1989.
[52] J. C. Hardwick and J. S. Lim, "The application of the IMBE speech coder to mobile communication," in *Proc. ICASSP*, May 1991, pp. 249–252.
[53] J. J. Y. Huang and P. M. Schultheiss, "Block quantization of correlated Gaussian random variables," *Trans. IRE*, vol. CS-11, pp. 289–296, 1963.
[54] D. Huffman, "A method for the construction of minimum redundancy codes," in *Proc. IRE*, Sept. 1952, pp. 1098–1101.
[55] ISO, "Coded representation of picture and audio information—Progressive bi-level image compression standard," ISO/IEC Draft, Dec. 1990.
[56] International Telephone and Telegraph Consultative Committee, "Facsimile coding schemes and coding control functions for group 4 facsimile apparatus," Red Book, Fasicle VII.3 Rec. T.6, 1980.
[57] A. E. Jacquin, "A novel fractal block-coding technique for digital images," in *Proc. ICASSP*, pp. 2225–2228, 1990.
[58] J. Jang and S. A. Rajala, "Segmentation-based image coding using fractals and the human visual system," in *Proc. ICASSP '90*, Apr. 1990, pp. 1957–1960.
[59] N. S. Jayant, "Adaptive quantization with a one-word memory," *Bell Syst. Tech. J.*, pp. 1119–1144, Sept. 1973.
[60] N. S. Jayant, *Waveform Quantization and Coding*. New York: IEEE Press, 1976.
[61] N. S. Jayant and S. W. Christensen, "Effects of packet losses on waveform coded speech and improvements due to an odd-even interpolation procedure," *IEEE Trans. Commun.*, vol. 29, pp. 101–109, Feb. 1981.
[62] N. S. Jayant and P. Noll, *Digital Coding of Waveforms: Principles and Applications to Speech and Video*. Englewood Cliffs, NJ: Prentice Hall, 1984.
[63] N. S. Jayant, J. D. Johnston, and Y. Shoham, "Coding of wideband speech," presented Proc. 2nd Europ. Conf. Speech Commun. Techol., Sept. 1991.
[64] N. S. Jayant, "High-quality coding of telephone speech and wideband audio," *IEEE Commun. Mag.*, pp. 10–19, Jan. 1990.
[65] J. D. Johnston, "A filter family designed for use in quadrature mirror filter banks," presented at Proc. ICASSP, 1980.
[66] ——, "Transform coding of audio signals using perceptual noise criteria," *IEEE J. Select. Areas Commun.*, pp. 314–323, Feb. 1988.
[67] B. H. Juang and A. H. Gray, "Multiple stage vector quantization for speech coding," in *Proc. ICASSP*, Apr. 1982, pp. 597–600.
[68] M. Kaneko, A. Koike, and Y. Hatori, "Coding with knowledge-

based analysis of motion pictures," in *Proc. Picture Coding Symp.*, June 1987, vol. 12-3, pp. 167-168.
[69] G. Karlsson and M. Vetterli, "Three dimensional subband coding of video," presented at Proc. ICASSP, 1988.
[70] ——, "Packet video and its integration into the network architecture," *IEEE J. Select. Areas Commun.*, pp. 739-751, June 1989.
[71] D. P. Kemp, R. A. Sueda, and T. E. Tremain, "An evaluation of 4800 b/s voice coders," presented at Proc. ICASSP, May 1989.
[72] T. Kim, "New finite state vector quantizers for images," in Proc. ICASSP, 1988, pp. 1180-1183.
[73] N. Kitawaki, M. Honda, and K. Itoh, "Speech-quality assessment methods for speech-coding systems," *IEEE Commun. Mag.*, pp. 26-32, Oct. 1984.
[74] N. Kitawaki and H. Nagabuchi, "Quality assessment of speech coding and speech synthesis systems," *IEEE Commun. Mag.*, pp. 36-44, Oct. 1988.
[75] N. Kitawaki, H. Nagabuchi, M. Taka, and K. Takahashi, "Speech coding technology for ATM networks," *IEEE Commun. Mag.*, pp. 21-27, Jan. 1990.
[76] W. B. Kleijn and W. Granzow, "Methods for waveform interpolation in speech coding," *Digit. Signal Proces.*, 1991.
[77] P. Kroon and B. S. Atal, "On improving the performance of pitch predictors in speech coding systems," in *Proc. ICASSP*, 1990, pp. 661-664.
[78] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-pulse excitation—A novel approach to effective and efficient multipulse coding of speech," *IEEE Trans. Signal Proces.*, vol. ASSP-34, no. 5, pp. 1054-1063, Oct. 1986.
[79] P. Kroon and K. Swaminathan, "A high quality multi-rate real-time CELP coder," this issue, pp. 850-857.
[80] M. Kunt, A. Ikonmopoulos, and M. Kocher, "Second-generation image coding technique," *Proc. IEEE*, pp. 549-574, Apr. 1985.
[81] D. LeGall, "MPEG: A video compression standard for multimedia applications," *Commun. ACM*, pp. 47-58, Apr. 1991.
[82] D. H. Lee and D. L. Neuhoff, "Conditionally corrected two-stage vector quantization," in *Proc. 1990 Conf. Inform. Sci. Syst.*, Princeton, NJ, Mar. 1990, pp. 802-806.
[83] ——, "An asymptotic analysis of two-stage vector quantization," presented at Proc. ISIT, 1991.
[84] Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantization design," *IEEE Trans. Commun.*, vol. 28, pp. 84-95, Jan. 1980.
[85] M. Liou, "Overview of the $p * 64$ kbits/s video coding standard," *Commun. ACM*, pp. 60-63, Apr. 1991.
[86] S. S. Magan, "Trends in DSP system design," in short course on *Digital Signal Processing*, IEEE-Int. Electron. Device Meet., Dec. 1989.
[87] J. Makhoul, S. Roucos, and H. Gish, "Vector quantization in speech coding," *Proc. IEEE*, pp. 1551-1588, Nov. 1985.
[88] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Patt. Anal. Mach. Intel.*, July 1989.
[89] J. Max, "Quantizing for minimum distortion," *IRE Trans. Inform. Theory*, pp. 7-12, Mar. 1960.
[90] R. J. McAulay and T. F. Quatieri, "Speech analysis and synthesis based on a sinusoidal model," *IEEE Trans. Signal Proces.*, pp. 744-754, Aug. 1986.
[91] P. Mermelstein, "G.722: A new CCITT coding standard for digital transmission of wideband audio signals," *IEEE Commun. Mag.*, pp. 8-15, Jan. 1988.
[92] N. Moayeri, D. L. Neuhoff, and W. E. Stark, "Fine-coarse vector quantization," *IEEE Trans. Inform. Theory*, pp. 1072-1084, July 1991.
[93] J. W. Modestino and D. G. Dant, "Combined source-channel coding of images," *IEEE Trans. Commun.*, pp. 1644-1659, Nov. 1979.
[94] S. Morishima, K. Aizawa, and H. Harashima, "A real-time facial action image synthesis system driven by speech and text," in *Proc. SPIE Vis. Commun. Image Proces.*, 1990, pp. 1151-1158.
[95] G. Morrison and D. Beaument, "Two-level video coding for ATM networks," *Sign. Proces.: Image Commun.*, pp. 179-195, June 1991.
[96] T. Murakami, K. Asai, and A. Itoh, "Vector quantization of color images," in *Proc. IEEE-ICASSP*, 1986, pp. 133-135.
[97] H. G. Musmann, "The ISO audio coding standard," presented at *Proc. IEEE GLOBECOM*, Dec. 1990.
[98] H. G. Musmann, P. Pirsch, and H.-J. Grallert, "Advances in picture coding," *Proc. IEEE*, pp. 523-548, Apr. 1985.
[99] N. M. Nasrabadi and R. A. King, "Image coding using vector quantization: A review," *IEEE Trans. Commun.*, pp. 957-971, Aug. 1988.
[100] A. N. Netravali and B. G. Haskell, *Digital Pictures*. New York: Plenum Press, 1988.
[101] A. N. Netravali and J. A. Stuller, "Motion compensation transform coding," *Bell Syst. Tech. J.*, pp. 1703-1718, Sept. 1974.
[102] A. N. Netravali and J. O. Limb, "Picture coding: A review," *Proc. IEEE*, pp. 366-406, Mar. 1980.
[103] A. N. Netravali, E. Petajan, S. Knauer, K. Mathews, R. J. Safranek, and P. Westerink, "A high quality digital HDTV codec," *IEEE Trans. Consum. Electron.*, pp. 320-330, Aug. 1991.
[104] D. L. Neuhoff and T. N. Pappas, "Perceptual coding of images for halftone display," presented at Proc. ICASSP, May 1991.
[105] D. L. Neuhoff and D. H. Lee, "On the performance of tree-structured vector quantization," presented at Proc. ICASSP, 1991.
[106] Y. Ninomiya, "HDTV broadcasting systems," *IEEE Commun. Mag.*, pp. 15-23, Aug. 1991.
[107] J.-R. Ohm and P. Noll, "Predictive tree encoding of still images with vector quantization," presented at Proc. ISSSE89, Nurmberg, Sept. 1989.
[108] J. B. O'Neal, "Predictive quantizing systems for the transmission of television signals," *Bell Syst. Tech. J.*, pp. 689-719, May/June 1966.
[109] K. S. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameters at 24 bits per frame," presented at Proc. ICASSP, 1991.
[110] T. N. Pappas and D. L. Neuhoff, "Model-based halftoning," presented at Proc. SPIE/IS&T Symp. Electron. Imag. Sci. Technol., Feb./Mar. 1991.
[111] J. Princen, A. Johnson, and A. Bradley, "Sub-band transform coding using filterbank designs based on time-domain aliasing cancellation," in *Proc. ICASSP*, 1987, pp. 2161-2164.
[112] C. I. Podilchuk, N. S. Jayant, and P. Noll, "Sparse codebooks for the quantization of non-dominant sub-bands in image coding," presented at Proc. ICASSP, 1990.
[113] W. Pratt, *Digital Image Processing*. New York: Wiley, 1978.
[114] S. R. Quackenbush, "Hardware impiementation of a color image decoder for remote database access," presented at Proc. ICASSP, 1990.
[115] ——, "A 7 kHz bandwidth, 32 kbps speech coder for ISDN," presented at Proc. ICASSP, 1991.
[116] L. R. Rabiner and R. W. Schafer, *Digital Speech Processing*. Englewood Cliffs, NJ: Prentice Hall, 1978.
[117] V. Ramamoorthy, N. S. Jayant, R. V. Cox, and M. M. Sondhi, "Enhancement of ADPCM speech coding with backward-adaptive algorithms for postfiltering and noise feedback," *IEEE J. Select. Areas Commun.*, pp. 364-382, Feb. 1988.
[118] B. Ramamurthi and A. Gersho, "Classified vector quantization of images," *IEEE Trans. Commun.*, pp. 1105-1115, Nov. 1986.
[119] A. R. Reibman, "DCT-based embedded coding for packet video," *Signal Proces.: Image Commun.*, pp. 333-343, Sept. 1991.
[120] P. Rioul and M. Vetterli, "Wavelets and signal processing," *IEEE Signal Proces. Mag.*, pp. 14-38, Oct. 1991.
[121] L. G. Roberts, "Picture coding using pseudo-random noise," *IRE Trans. Inform. Theory*, pp. 145-154, Feb. 1962.
[122] G. Roy and P. Kabal, "Wideband CELP speech coding at 16 kbps," in *Proc. ICASSP*, 1991, pp. 17-20.
[123] M. J. Sabin and R. M. Gray, "Product vector quantizers for waveform and voice coding," *IEEE Trans. Signal Proces.*, pp. 474-488, June 1984.
[124] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image-dependent quantization and post-quantization," in Proc. ICASSP, 1989.
[125] T. Saito, H. Takeo, K. Aizawa, H. Harashima, and H. Miyakawa, "Adaptive image coding using gain-shape vector quantization," in *Proc. IEEE-ICASSP*, Apr. 1986, pp. 129-132.
[126] D. J. Sakrison, "Image coding applications of vision models," in *Image Transmission Techniques*, W. K. Pratt, Ed. New York: Academic, May 1979, pp. 21-51.
[127] W. F. Schreiber, "Psychophysics and the improvement of television picture quality," *SMPTE J.*, pp. 717-725, Aug. 1984.
[128] J. Schroeter and M. M. Sondhi, "Speech coding based on physiological models of speech production," in *Advances in Speech Signal*

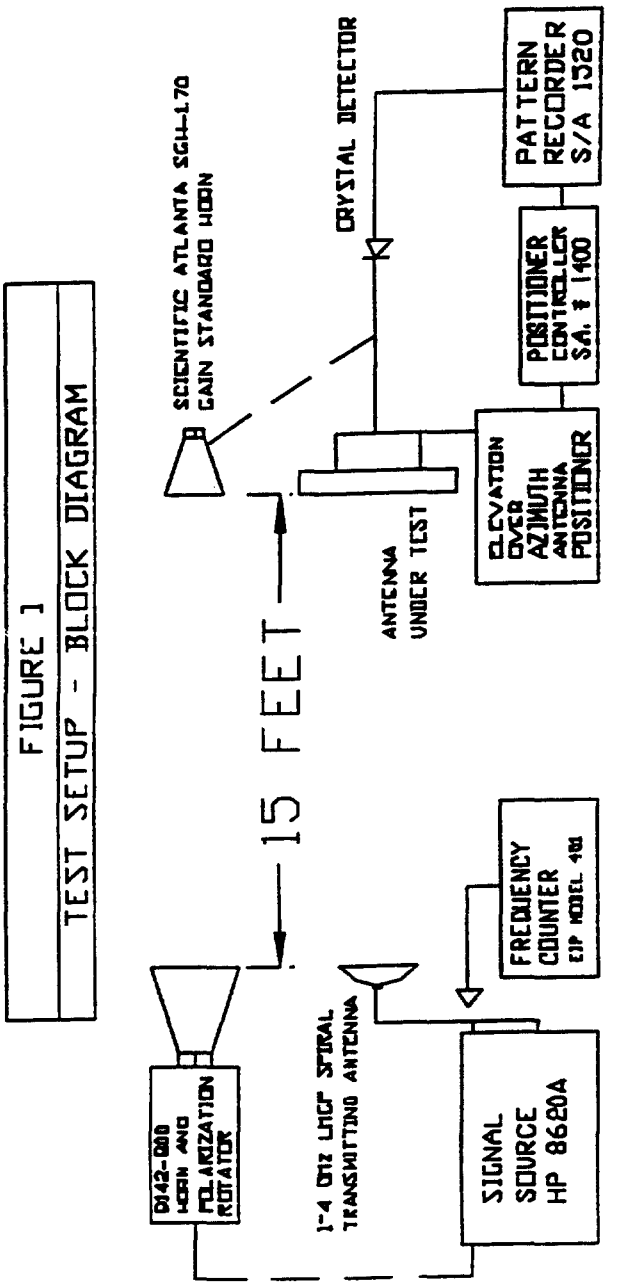
Processing, S. Furui and M. M. Sondhi, Ed. New York: Marcel Dekker, 1991.
[129] C. E. Shannon, "A mathematical theory of communications," *Bell Syst. Tech. J.*, vol. 27, 1948, pp. 379-423, and 623-656.
[130] ——, "Coding theorems for a discrete source with a fidelity criterion," *IRE Nat. Conv. Rec.*, part 4, pp. 142-163, 1959.
[131] Y. Shoham, "Constrained excitation coding of speech at 4.8 kbps," in *Advances in Speech Coding*. New York: Kluwer Academic, 1991, pp. 339-348.
[132] B. Smith, "Instantaneous companding of quantized signals," *Bell Syst. Tech. J.*, pp. 653-709, May 1957.
[133] R. Steele, *Delta Modulation Systems*. New York: Halsted, 1975.
[134] ——, "The cellular environment of lightweight handheld portables," *IEEE Commun. Mag.*, pp. 20-29, July 1989.
[135] T. G. Stockham, "Image processing in the context of a visual model," *Proc. IEEE*, pp. 828-842, July 1972.
[136] Swedish Radio, unpublished report on the quality of low rate audio algorithms submitted for the ISO-MPEG standard, Aug. 1990.
[137] G. Theile, G. Stoll, and M. Link, "Low bit-rate coding of high-quality audio signals," *EBU Tech. Rev.*, no. 230, pp. 71-94, Aug. 1988.
[138] H. Tominaga, H. Jozawa, M. Kawashima, and T. Hanamura, "A video coding method considering cell losses in ATM networks," *Signal Process.: Image Commun.*, pp. 291-300, Sept. 1991.
[139] T. E. Tremain, "The government standard linear predictive coding algorithm: LPC-10," *Speech Techno.*, vol. 1, no. 2, pp. 40-49, Apr. 1982.
[140] P. P. Vaidyanathan, "Quadrature mirror filter banks, M-band extensions and perfect-reconstruction techniques," *IEEE ASSP Mag.*, pp. 4-20, 1987.
[141] J. Vaisey and A. Gersho, "Variable block-size image coding," in *Proc. ICASSP*, Apr. 1987, pp. 1051-1054.
[142] R. Valenzuela and C. N. Animalu, "A new voice-packet reconstruction technique," in *Proc. ICASSP*, 1989, pp. 1334-1337.
[143] W. Verbiest and L. Pinnoo, "A variable bit rate video coder for asynchronous transfer mode networks," *IEEE J. Select. Areas Commun.*, pp. 761-770, June 1989.
[144] M. Vetterli, "Multidimensional sub-band coding: Some theory and algorithms," *Signal Proces.*, pp. 97-112, Apr. 1984.
[145] W. D. Voiers, "Diagnostic acceptability measure for speech communication systems," in *Proc. ICASSP*, May 1977, pp. 204-207.
[146] ——, "Diagnostic evaluation of speech intelligibility," in *Speech Intelligibility and Speaker Recognition*, M. Hawley, Ed. Stroudsburg, PA: Dowden Hutchinson Ross, 1977.
[147] G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, pp. 31-43, Apr. 1991.
[148] S. Wang and A. Gersho, "Phonetically-based vector excitation coding of speech at 3.6 kbps," in *Proc. ICASSP*, May 1989, pp. 49-52.
[149] L. Wang and M. Goldberg, "Progressive image transmission using vector quantization on images in pyramid form," *IEEE Trans. Commun.*, pp. 1339-1349, Dec. 1989.
[150] P. H. Westerink, D. E. Boekee, J. Biemond, and J. H. Woods, "Subband coding of images using vector quantization," *IEEE Trans. Commun.*, pp. 713-719, June 1988.
[151] R. Wilson, H. 'E. Knutsson, and G. H. Granlund, "Anisotropic nonstationary image estimation and its applications: Part II—Predictive image coding," *IEEE Trans. Commun.*, pp. 398-406, Mar. 1983.
[152] P. A. Wintz, "Transform picture coding," *Proc. IEEE*, pp. 809-820, July 1972.
[153] I. H. Witten, R. M. Neal, and J. G. Cleary, "Arithmetic coding for data compression," *Commun. ACM*, pp. 520-540, June 1987.
[154] J. W. Woods and S. D. O'Neil, "Subband coding of images," *IEEE Trans. Acoust., Speech, Signal Proces.*, pp. 1278-1288, Oct. 1986.
[155] S.-W. Wu and A. Gersho, "Optimal block-adaptive image coding with constrained bit rate," presented at Proc. 24th Asilomar Conf. Signals, Syst. Comput., Nov. 1990.
[156] A. D. Wyner, "Fundamental limits in information theory," *Proc. IEEE*, pp. 239-251, Feb. 1981.
[157] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," *IEEE Trans. Acoust., Speech, Signal Proces.*, pp. 299-309, Aug. 1977.

Nikil Jayant (M'69-SM'77-F'82) received the Ph.D. degree in electrical communication engineering from the Indian Institute of Science, Bangalore, India.

He is Head of the Signal Processing Research Department at AT&T Bell Laboratories in Murray Hill, NJ. He is responsible for research in speech and image processing with applications to coding, communications, and recognition. He joined AT&T Bell Laboratories in 1968. He is the Editor of *Waveform Quantization and Coding* (New York: IEEE Press, 1976) and coauthor of *Digital Coding of Waveforms—Principles and Applications to Speech and Video* (Englewood Cliffs, NJ: Prentice-Hall, 1984).

Dr. Jayant was the first Editor-in-Chief of the IEEE ACOUSTICS, SPEECH, AND SIGNAL PROCESSING MAGAZINE.

Figure A3-8 Antenna Test Set-Up



was place at the other end of the chamber to be used as a reference receiver. The gain of the AUT is then determined by comparing the gain of the second SGT antenna to that of the AUT.

Axial ratio was taken by rotating a SGT antenna - Model #9142-800 (source antenna) at one end of the chamber while the AUT receives the signal at the other end of the chamber. The amplitude difference of the linear components is then recorder by the instrumentation receiver and rectangular chart recorder. Axial ratio was taken in the elevation plane. Swept return loss was recorded using a Hewlett Packard Model 8350B Sweep Oscillator and a Model 8756A Scalar Network Analyzer. The results of the antenna measurements show that the desired performance was generally achieved. The results are presented in Section 4 which include constraints imposed by the experiment emulation physical range.

The antenna built for the experiment was constrained by the requirement to operate at 10° elevation angle rather than the 20° minimum elevation angle when in operation with the planned satellites. The 10° elevation angle is required in the Demonstration due to the limited height of the buildings on which the S-band transmitters were located. The main effects were lower gain, somewhat higher gain ripple and lower effective transmission bandwidths.

A3.5 MOBILE VEHICLE

The mobile vehicle signal path starts at the small diameter S-band antenna embedded into the vehicle rooftop. This circularly polarized antenna offers an omni-directional pattern in azimuth and an elevation beamwidth of 40° centered at a nominal elevation of 35° with 3 dB peak antenna gain. The remaining RF chain as illustrated in Figure A3-8 consists of an LNA mounted under the antenna, a second gain stage to buffer/compensate for cable losses, a down-converter circuit to bring the S-band signal into the L-band region required by the satellite modems and a 4-port power splitter to drive the four frequency diverse satellite modems.

As illustrated in Figure A3-9, the remaining signal processing elements are

Figure A3-8 Automobile RF Receiver Equipment

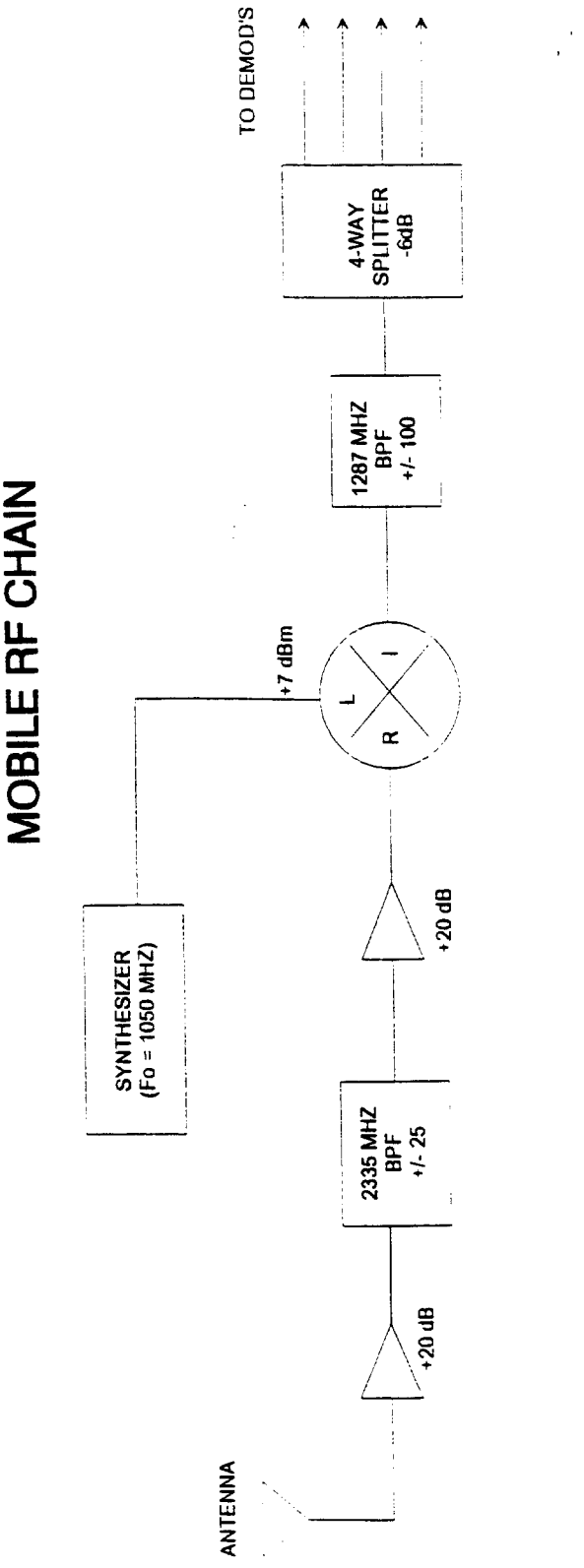
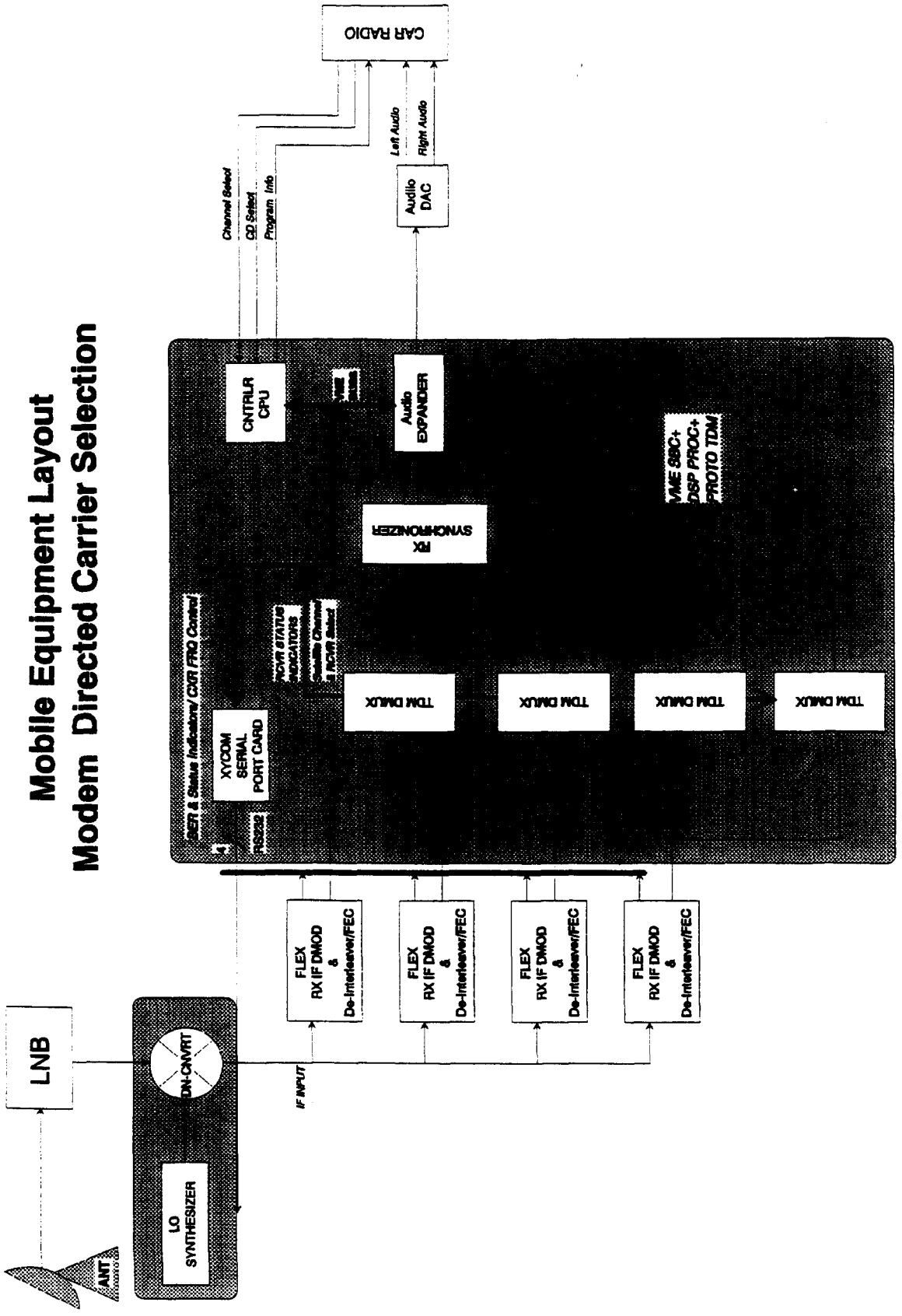


Figure A3-9 Automobile Signal Processing Equipment



concerned with the baseband signal from the demodulators and theregeneration of the transmitted CD quality source material..

The down converted L-band RF signal is first distributed to the four (4) OQPSK demodulators, each tuned to one of the four carriers used in the test range (2322, 2330, 2338, 2346 MHz). In addition to carrier demodulation, the demodulators then lock to the interleave framing and de-interleave the received data stream and pass the result through a monolithic Viterbi convolutional FEC decoder. The demodulated, de-interleaved and FEC decoded/corrected output data is then sent to their separate TDM demultiplexers.

In addition to their output data ports, each demodulator is equipped with a maintenance and control port which is used to pass performance data to a 68000 VME based single board computer (SBC) which serves as an overall executive or receiver controller for the mobile receiver unit and as the interface for the radio control/display unit. In the receiver controller role, the SBC uses a four serial port VME slave card to monitor the performance of all 4 demodulators and based on the results of this monitor, it selects one of the four Demod/DMUX chains as the "primary" or "master" circuit.

The primary DMUX circuit is then used to supply the demodulated compressed audio data to the Audio decoder/expander along with its associated clock and octet frame pulse. This octet/byte frame pulse is also used, along with a Multi-frame pulse, by the slave DMUX circuits to track the output phase or timing of the master circuit and thus ensure synchronization and hitless switchover at the input of the audio decoder.

In addition to audio data, the primary DMUX stores all of the received NCC data in a local FIFO which may then be read by the 68000 SBC and processed to extract the relevant part of the NCC for the currently selected radio channel. This radio information data is then sent by the SCC over a fifth serial port to the radio/display controller.

It should be noted that the switchover performance between

Demod/DMUX chains (i.e., between S-band carriers) is critical to the planned strategy of dual satellite continental coverage. To this end, every effort has been made to ensure a "hitless" switchover between DMUX circuits and this is the role of the RX synchronizer. In fact, the RX synchronizer hardware is actually distributed over the four DMUX circuits. At any instant of time there is one "MASTER" circuit which is on-line and delivering its communication payload to the audio decoder and the SBC (i.e., the NCC data) and three "SLAVE" circuits which are simply tracking the output phase/byte-number of the output multi-frame data from the MASTER circuit.

The seamless switchover requirement then essentially involves ensuring that switchover occurs at byte/octet boundaries. There are no extra clock edges during switchover, all slave DMUX circuits track the primary/master output timing to ensure that they can output the next consecutive byte into the multi-frame should they become the next "master"

Under this design constraint the only effect of switchover is a delay/step change in phase of the audio data clock phase in order to pick up the new MASTER's active clock edge.

The DMUX circuits are implemented on standard VME prototype cards with two circuits per VME card. As these cards employ wire-wrapped digital circuits they are necessarily 2 units wide. Circuit density, flexibility and rapid prototyping was achieved by using MACH230/130 type PLD technology for all logic design components with the exception of the line interface components.

The compressed 128 Kbs audio data streams from the DMUX drives a VME DSP32C based real-time decoder platform. The audio decoder/expander is responsible for restoring the original 48 KHz sample rate audio data rate, formating and sending this over a standard AES serial data stream (approximately 1.54 Mb/s) to professional stereo CD quality reference DACs, APOGEE Electronics Model DA-1000. The analog signal from the DACs is then routed through a relay switch to the analog

"CD" input of the radio for audio playback.

ATTACHMENT A4. MUSIC PROGRAMMING

APPENDIX B

CD RADIO DARS TECHNOLOGY INNOVATIONS

SUMMARY

CD Radio has conceived, pioneered, developed and demonstrated several technology innovations over the 1990-1993 time frame which permit the economical provision of Digital Audio Radio Service (DARS) by geosynchronous satellite transmission to mobile vehicles and other users.

INTRODUCTION

The provision of DARS by geosynchronous satellite transmission, both nationally and internationally and for both mobile and stationary users, has been proposed and actively studied since 1980. CCIR Report 955-2 "Satellite Sound Broadcasting To Vehicular, Portable And Fixed Receivers In The Range 500-3000 MHz" (Reference 1) well summarizes the past theoretical and analytical work. In 1990, CD Radio embarked on an intensive technical effort to design a geosynchronous satellite based DARS for users in the United States. The user population was determined early in the design effort to be primarily mobile vehicles (e.g., automobiles). It was also determined early that the CD Radio DARS would provide many, very high audio quality music channels on a subscription basis nationwide which should clearly differentiate the service by users from terrestrial local radio.

The CD Radio design effort extended over a few years and, initially, it was found impossible to provide the required service previously mentioned in an economical

manner. First, and most important, DAR service to mobile vehicles in the United States from a single, high power geosynchronous satellite would likely be unsatisfactory due to blockage. Essentially mobile vehicles going under the numerous highway overpasses and having the line of sight from the satellite to the mobile vehicles interrupted by trees, buildings and large trucks in a suburban/urban physical environment would create numerous enough short service outages to be objectionable to listeners. The number, frequency and length of such outages are dependent on many factors including vehicle speed of motion and elevation angle from the vehicle to the satellite.

Second, such a design would require large and expensive satellites. As proposed by others studying such designs and well documented in Reference 1, the satellite Effective Isotropic Radiated Power (EIRP) required for continental Unitd States service would be extremely large, necessitating enormous satellite prime power supplies and batteries. It is shown in Reference 1 and Reference 2 that satellite beam edge EIRPs approaching 70 dBW are needed for such DAR service requiring a satellite radio transmitter total power of at least 20,000 watts and a prime power supply of 50,000 watts. Over 90% of this power is used to overcome multipath fading. Such a very high powered geosynchronous satellite is believed to be somewhat beyond the current commercial state-of-the-art but, without arguing that aspect, such a satellite is unarguably expensive. Such a satellite, its development and launch is believed to cost over $500M without consideration of reliability and backup facilities.

CD RADIO'S INNOVATIONS

CD Radio has designed a practical DARS using geosynchronous satellite transmissions by original innovations in technology, by innovative combinations

of existing technology and by sponsoring adaptation of emerging technology being developed by other organizations. These innovations are:

1. Satellite Spatial Diversity
2. Satellite Radio Frequency Diversity
3. Satellite System Technology (i.e., 128 kb/s CD music compression, automobile radio design and silver dollar sized mobile vehicle antennas)

Satellite Spatial Diversity

The first CD Radio technical innovation is to broadcast simultaneously the same digital audio radio signal from two geosynchronous satellites widely separated on the geosynchronous orbital arc but located so that all mobile users in the coverage area have an elevation angle to both satellites greater than 20° and have widely different azimuth angles. Such a satellite DARS system design results in very significant system performance improvements described below:

1. Blockage. Essentially the vehicle receiver is capable of receiving the transmission from both satellites and either selects the strongest one or combines them. Since each satellite transmission is above 20° in elevation angle and at a different azimuth angle, the probability of an outage due to blockage is greatly diminished over a single satellite DARS system. This is graphically shown by Figure 1, and Figure 2 shows the elevation angles for representative locations. Note that even in the lowest case shown of 23.3° elevation angle (Portland, Oregon user vehicle to the 80° West Longitude satellite), the user vehicle has an elevation angle of 36.6° to the other satellite.

2. Multipath. Satellite spatial diversity creates a large improvement in both Ricean and Rayleigh non-frequency selective transmission amplitude fading. This stems from the fact that fading is time and location dependent (Reference 3) and the fading on two spatially independent transmission paths will be statistically uncorrelated, so the probability of having both paths fade simultaneously will be greatly reduced. This improvement by spatial diversity has been proved by Bell Laboratories a few decades ago on terrestrial microwave radio relay paths where multipath is much worse due to the 0-5° path elevation angles typically employed. It is estimated that the transmission link margin allocated for multipath can be reduced at least eightfold (i.e., 9 dB) over the margin which would be allocated without satellite spatial diversity. Fading due to frequency selective multipath effects is considered on the following page.

3. Satellites. The improvements noted in the foregoing two paragraphs require the satellite DARS system to employ two in-orbit satellites. Although it might appear this is more expensive, the satellite spatial diversity system is much more economical than a single satellite DARS system. This is because each satellite in the diversity satellite DARS system requires one-eighth the EIRP (and prime power) for equivalent service. The CD Radio DARS system using dual satellites for spatial diversity will cost $250M including launch vehicles, which is half that earlier estimated for the high power single satellite DARS system. Besides the savings of $250M, the satellite spatial diversity system also substantially reduces outages from blockage. Equally important, the reduction in required satellite EIRP by the previously described CD Radio innovation allows a lower satellite power flux density along the United States-Canadian border which is necessary to achieve interference coordination of satellite DARS with Canada.